Validation of the Dutch Airforce test battery using neural networks

Markus Sommer
Joachim Häusler
Marco Koning
Martin Arendasy

info@schuhfried.at
www.schuhfried.at

Problem of criterion validation

- In most aviation psychological validation studies the magnitude of the correlation coefficients between an individual tests and the criterion measure is often low (c.f. Hunter & Burke, 1994; Burke et al., 1997).

- **Causes:**
  - Low reliability of predictor and criterion measure (Goeters,1998)
  - Attenuation of the variance in the predictor measure (Goeters, 1998)
  - Lack of symmetry between predictors and criterion (Wittman & Süß, 1997)

Lack of symmetry

Predictor variables:

Modular measure

Preferable one-dimensional

Simplified performance samples

Criterion variable:

Unified global measure

Multicausally determined

Complex performance samples

The magnitude of the correlation coefficient between a single predictor variable and a global criterion measure is necessarily low.

→ **Solution:** Validation of the test battery

Methods of criterion validation

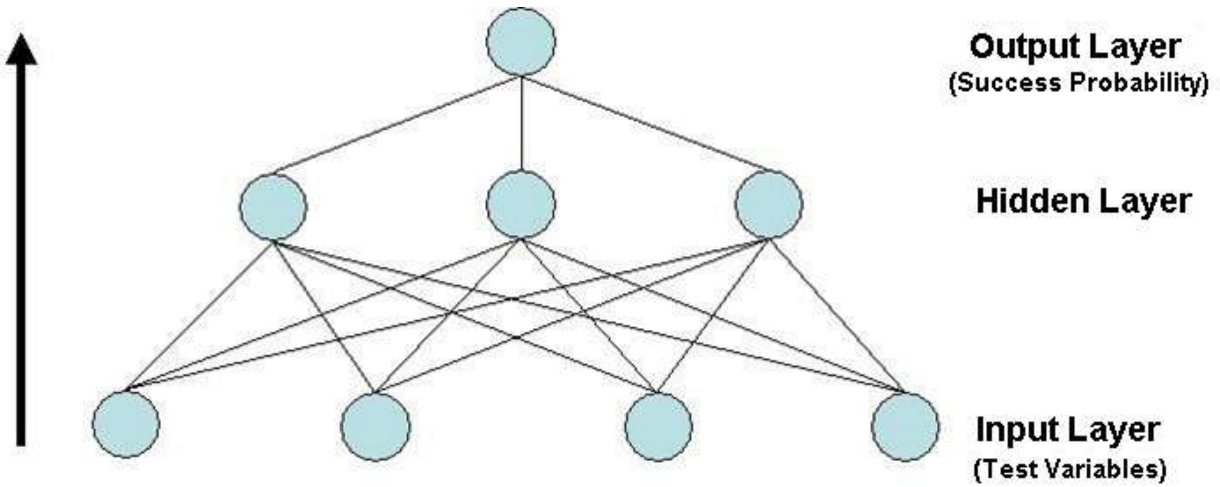
- **Most important methods:**
  - Discriminant analysis, multiple regression and logistic regression

- **Problems of these classical methods:**
  - Vulnerability to violations of their assumptions (c.f. Brown & Wicker, 2000).
  - Non-linear relations cannot be taken into account (c.f. Rakes, 1991).
  - Low stability in cross-validation studies (c.f. Jahnke, 1982).

Artificial neural network

- Artificial neural networks are systems of equations designed for pattern recognition tasks (c.f. Bishop, 1995; Kinnebrock, 1992; Rojas, 2000).

- Artificial neural networks make little assumptions about data characteristics.



Construction of an Artificial Neural Network

- Determination of the **architecture** of the artificial neural network

- Determination of the **number of predictor variables** and **hidden layer units** using parsimony indices (c.f. BIC) and adj. R^2 to avoid an over-fit of the model (Häusler & Sommer, 2005).

- **Optimization of the model weights** in the learning sample by minimizing the difference between actual criterion values and predictions.

- **Assessing the stability** of the artificial neural network solution using (1) leave one out validation, (2) Bootstrap-validation and (3) cross-validation.

Previous applications of artificial neural networks

- In previous studies artificial neural networks proved to be a valid and useful tool in traffic psychological assessment as well as occupational psychology (c.f. Sommer, Arendasy, Schuhfried & Litzenberger, 2005; Häusler & Sommer, 2005).

- Griffin (1998) evaluated artificial neural networks using a test battery primary consisting of psychomotor tests. Artificial neural networks resulted in higher validity coefficients than classical methods of statistical judgment formation but the difference did not reach statistical significance.

- Sommer, Olbrich and Arendasy (2004) used a broader test battery than the one used by Griffin and found higher classification rates and validity coefficients using artificial neural networks The difference to the results obtained by Griffin (1998) are attributed to differences in the breath of the chosen test battery in the two studies.

Methode

Predictor variables:

The following tests were taken from the Expert System Aviation (Schuhfried, 2005):

- Numerical-inductive reasoning (NID)
- Figural-inductive reasoning (FID)
- Computational Estimation (ASF)
- Arithmetic Competence (AK)
- Numerical Flexibility (ANF)
- Focused attention (COG)
- Mental rotation (A3DW)
- Decision speed (EF)
- Inspection time (BZ)

Criterion measure:

Global evaluation of the applicants‘ performance in the standardized flight simulator.

Sample

Military Pilot Applicants (N=99):

Sex: 99 % male; 1 % female

Age: 16 to 25 years (Mean=18.84, SD=2.04)

A total of 38 (38.4 %) respondents received a positive global evaluation of their performance in the standardized flight simulator.

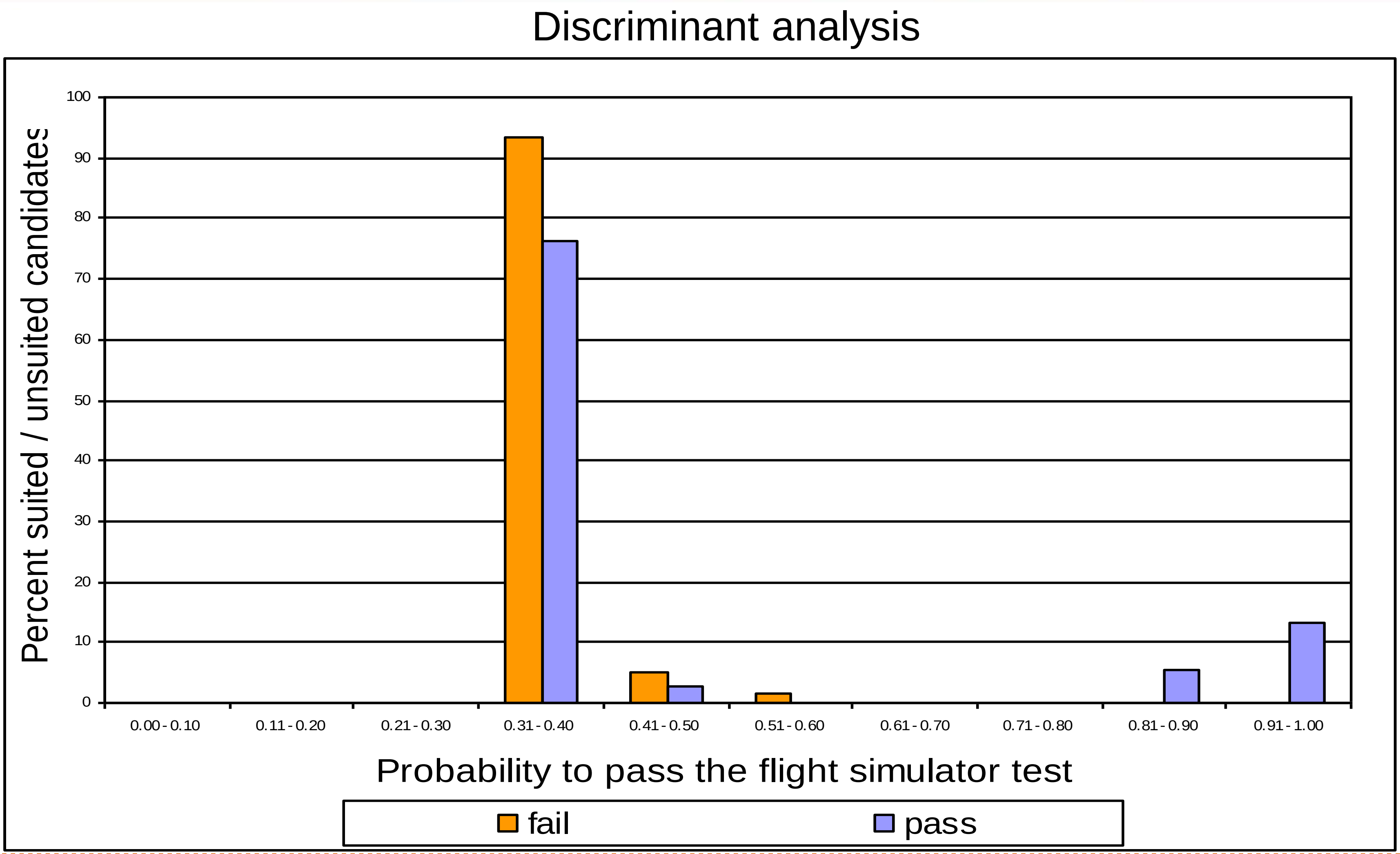
Results: Model optimization

Predictor variables	Hidden layer elements	BIC	adj R^2
NID, ASF, AK, EF, NF, A3DW	2	351.70	.256
NID, AK, BZ, EF, FID, TAVT	3	352.81	.512
ASF, AK, BZ, EF, FID, TAVT	1	357.91	.114
ASF, AK, EF, FID, TAVT	6	368.45	.790
NID, ASF, AK, BZ, FID, A3DW	5	373.96	.663
NID, AK, BZ, EF, FID, NF, TAVT	**3**	**345.22**	**.599**
ASF, AK, NF, BZ, EF, FID, TAVT	5	373.13	.750

Results

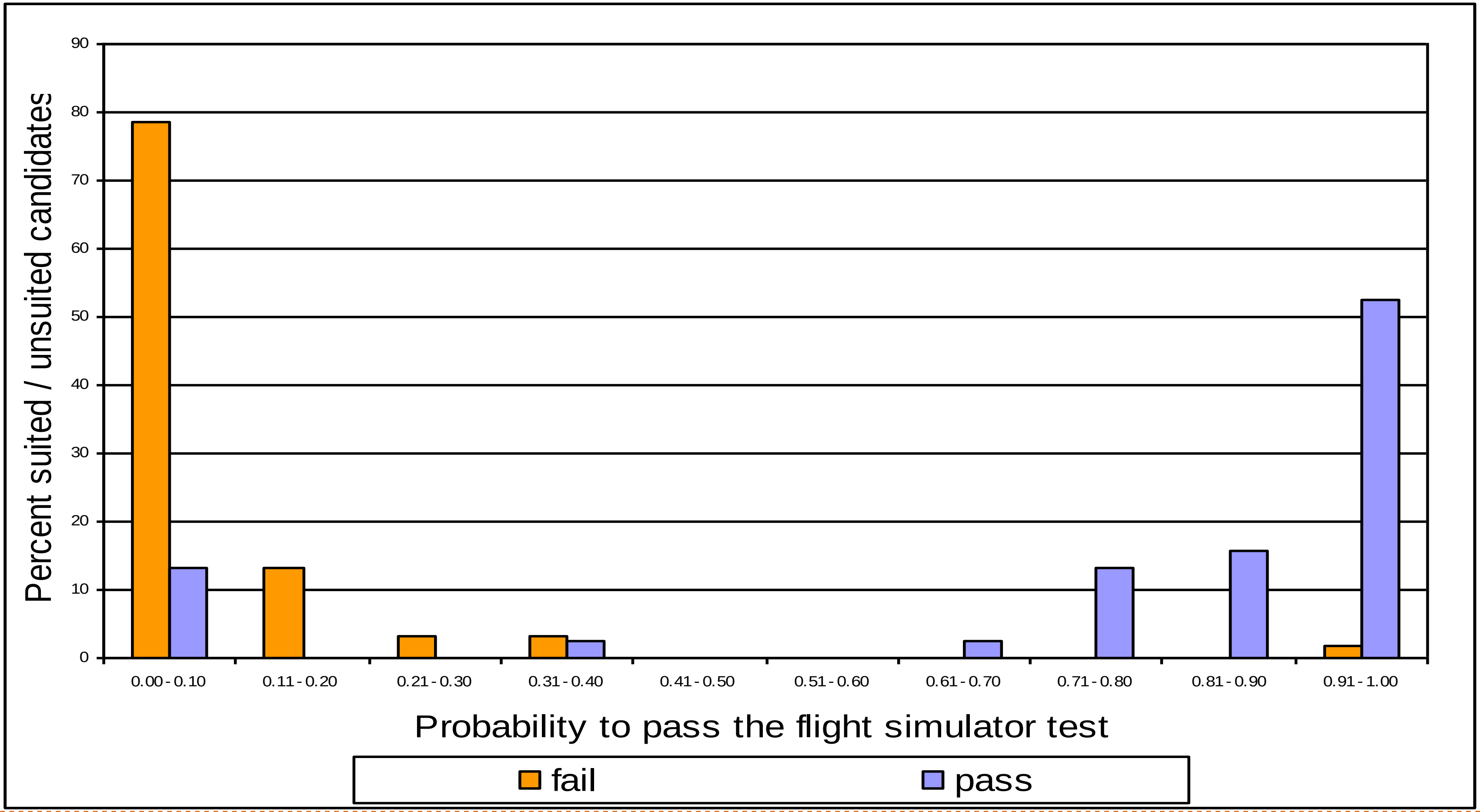
Discriminant analysis	Artificial neural network
Total sample (N=99) Box-M: F = 1.421; p = .069 Wilks Lambda = .93; df = 7; p = .418 **Classification rate: 58.6 %** **Validity coefficient: .26**	**Total sample (N=99)** Hidden layer units: 3 Quickprop; Iterations: 5000 **Classification rate: 92.9 %** **Validity coefficient: .84**
Jackknife Validation (N=99) **Classification rate: 48.5 %** **Validity coefficient: .20**	**Jackknife Validation (N=99)** **Classification rate: 92.9 %** **Validity coefficient: .83**
Bootstrap Validation Bootstrap samples: 1000 **Classification rate: [53.0%;73.4%]** **Validity coefficient: [.12; .52]**	**Bootstrap Validation** Bootstrap samples: 1000 **Classification rate: [88.2%; 97.7%]** **Validity coefficient: [.74; .94]**

Classification probability



Classification probability

Artificial neural network



Predictor relevance

dimension	incremental validity	relative relevance
Numerical-inductive reasoning (NID)	.157	13.4 %
Arithmetical competence (AK)	.205	16.9 %
Inspection time (BZ)	.195	16.2 %
Decision quality and speed (EF)	.167	14.1 %
Figural-inductive reasoning (FID)	.324	24.6 %
Numerical flexibility (ANF)	.105	4.6 %
Overview (TAVT)	.117	10.2 %

Discussion

- Artificial neural networks produced stable results and outperformed classical multivariate methods. This is in line with previous research (c.f. Häusler & Sommer, 2005; Sommer et al., 2005).

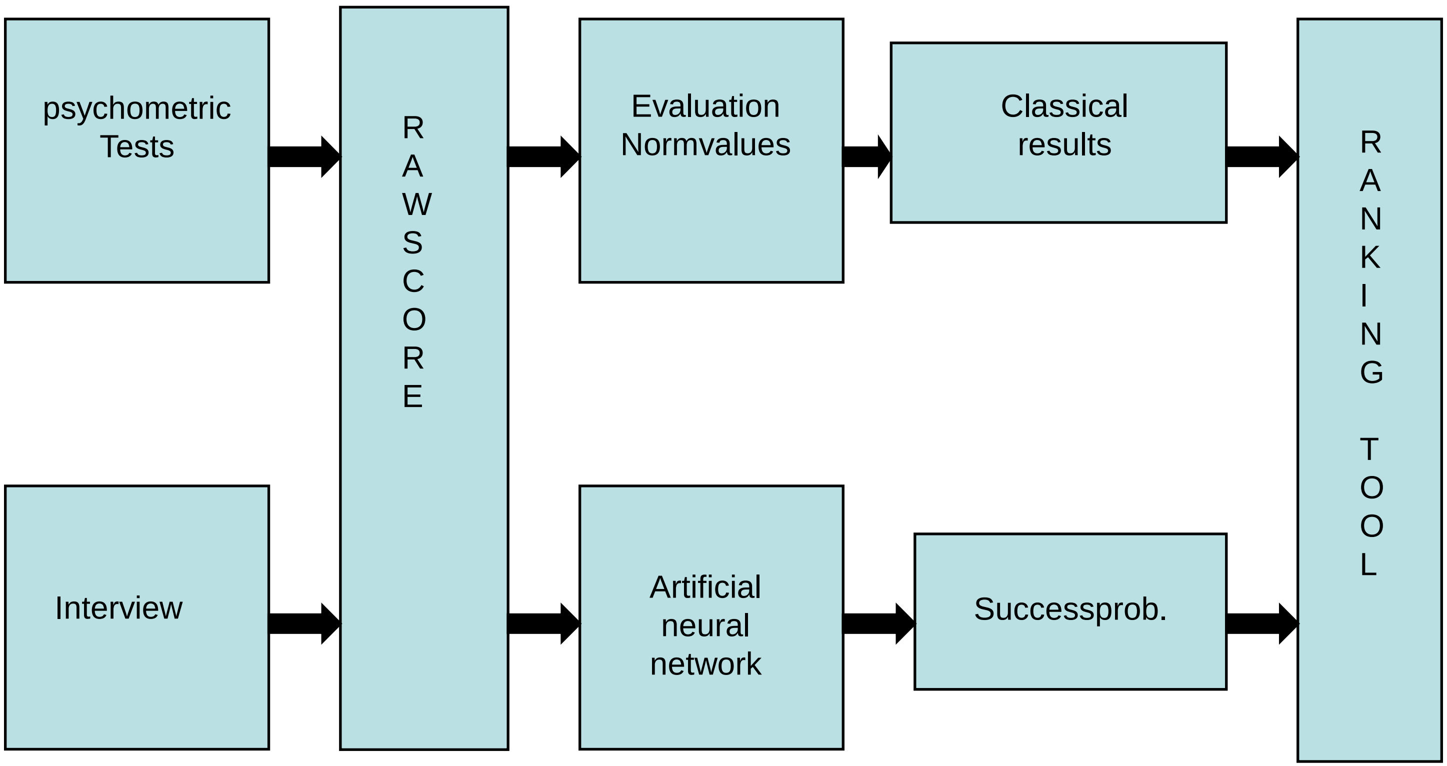
- Artificial neural networks also outperform classical multivariate methods regarding the separability of successful and less successful military pilot applicants.

- The results indicate, that measures of fluid intelligence (G_f), quantitative reasoning (G_q) and mental speed (G_s) contribute the most to predictive model.

- Mental rotation (G_v) did not provide any incremental validity in the present study.

- The authors acknowledge, that the results should be cross-validated using a second independent data set.

Practical application of the artificial neural network



The artificial neural network calculates a success probability for each candidate.